

# Notes on Databases:

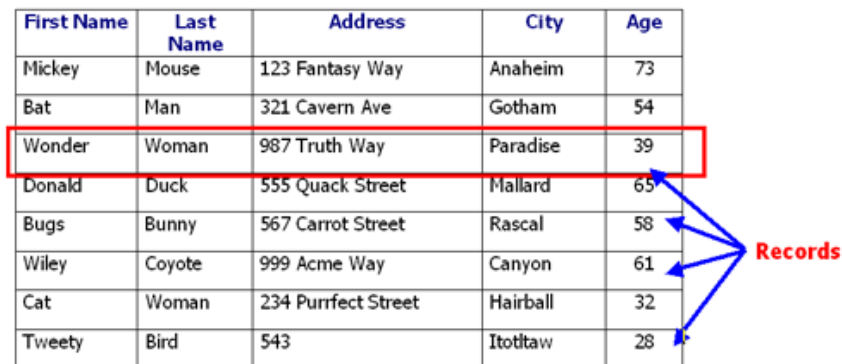
## A little background information

- Databases store data or information in tables, just like the one below:

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28

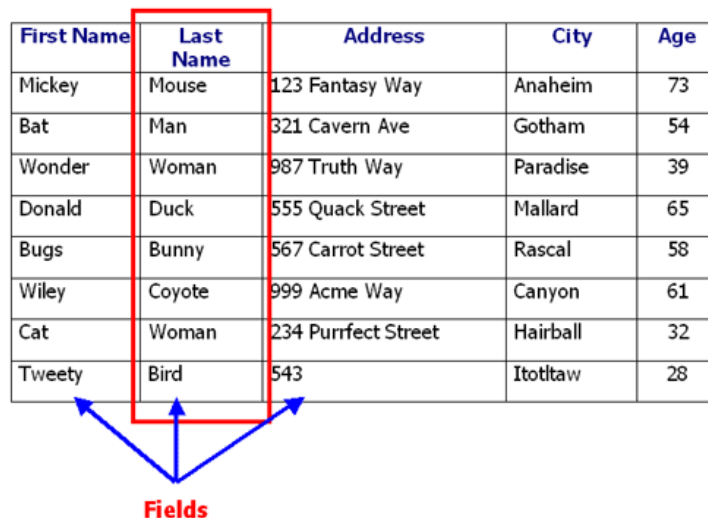
- Each table contains a lot of records:
  - o **DEFINITION:** A record is all of the data or information about one person or one thing.
  - o In the table below, all of the information about each cartoon character is stored in a 'row' or 'record'.

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28



- A record is made up of lots of individual pieces of information. Look at Wonder Woman's record: it stores her first name, last name, address, city and age.
  - o Each of these individual pieces of information in a record is called a 'field'.
  - o **DEFINITION:** a 'field' is one piece of data or information about a person or thing.

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28



- **Two types of databases**
  - Flat file – when all of the data is held in one big table
    - Can create data duplication
  - Relational databases
    - The data is split up into sensible groups then a separate table is made for each group.
    - The main benefit of a relational database is that data doesn't have to be duplicated.
    - Reducing data duplication reduces the amount of data which needs to be stored, thus making the database smaller. It also reduces the risk of mistakes, because every time you have to type the same data in, there is a risk you could miss-spell it.
  - **ITGS Question?** Identify three tables that could be part of a doctor's office patient relational database.
 

**Answer** – Table of patient's names, table of patient's medicines, table of patients known allergies, table of patient's weight, table of patient's date of visits
- **Queries**
  - Searching the database to find specific information is known as a query.
  - Data mining is the process of using queries to find specific information and relationships in a database
  - Queries can be made specific to the type of data in a field and by indicating less than/more than or equal to.
  - **ITGS Question?** A librarian needs to produce a list of drama books the Year 12 students have not returned on time. Describe the query needed.
 

Type=drama  
Due date < today's date or Returned=no  
Year=12
  - **ITGS Question?** A doctor needs to produce a list of patients who have diabetes and has been prescribed "drug a" and "drug c" in the past six months. Describe the query needed.
 

Patient condition = diabetes  
Drugs prescribed="drug a", "drug c"  
Prescription date for "drug a" ≥ six months ago  
Prescription date for "drug c" ≥ six months ago

## Why are we concerned about electronic databases?

- **Stores Personal Information**
  - SSN, Account information, Medical Information, order information, etc...
- **Why is this important?**
  - Identity Theft
    - Personal information stolen and used to secure credit cards, SSN cards and other personal identification in someone else's name for financial gain.
  - Medical Misdiagnosis
    - Improperly stored information could result in a misdiagnosis, increases the chance of receiving wrong medicine.
  - Orders and/or reservations might be messed up causing problems
    - Receiving the wrong product purchased
    - Wrong plane, hotel, movie, rental car, etc... reserved
    - Not receiving product purchased or getting reservation

- **What do we need to protect electronic databases from?**

- External hacking
  - Someone outside of the company/organization network illegally gains access to the network and databases therein.
    - How to protect against?
      - **Firewalls** – Have one properly configured for your network
        - Can be either hardware or software based
        - Can be configured to only allow access to enter and/exit a network by certain websites, file downloads, certain ports on the computers, particular IP Addresses, specific computers, specific users
      - **Anti-virus/anti-spyware** – Use anti-virus and anti-spyware software to protect against malware. Keep it up-to-date, always download and install critical updates for your OS and other computer software.
        - Any type of computer infection: viruses, Trojan horses, spyware, worms, logic bombs, etc...
        - Can enter a computer network via many different ways, such as, emails, email attachments, website links, in free and benign looking software, via storage media brought in by network users like floppy discs, USB keys, CD's and DVD's, instant messaging programs and other ways...
          - Infected email attachments can be in any type of file including, .exe, .doc, and .pdf
        - Can be written to track computer user's Internet usage, retrieve saved passwords and form data, record users' keystrokes, damage registry files or other software files on the computer, etc...
      - **Encryption** – encrypt data so only those authorized to view it can with proper de-encryption software.
- Unauthorized access to database within
  - Someone within or outside of the organization gets access to data they are not supposed to have access to
    - How to protect against?
      - Strict company/organizational policies regarding who has access to information and who does not.
      - Authentication technology used to access data
        - Passwords/usernames
        - Security questions
        - Biometrics
          - Retina scans
          - Fingerprints
          - DNA
          - Facial recognition
          - Voice recognition
        - RFID tags
          - Uses radio frequency to transmit an ID number to a scanner. The ID number is matched to a database record

- Smart Cards
    - Small amount of data identifying user is stored on the card
  - Bar Codes
    - Magnetic ID strip is scanned and matched to a database record
- CAPTCHA Systems - using a puzzle to identify that the user is a human, not a bot attack (Commonly known as images with a numerical/letter combination on an image and the user is asked to identify the numerical/letter combination)
- Data is released to a third party without consent
  - Business or organization must have a policy that addresses third party notification
    - Needs to clearly notify those in the database if their information is going to be released to a third party
    - Consent must be given by those in the database before release and an “opt-out” option should be given
- Ownership of company or organization changes, what happens to data?
  - Current policy needed to address what happens to data if the company or organization is sold
    - If new policies are instituted by new owner, they need to be presented to customers and customers would have the ability to choose if they want to continue to be a part of their clientele, if they choose not to, then data about them needs to be removed and verified by new owner
- Dirty data
  - Incorrect record information in the database
    - Business/organization needs to make sure that the database does not contain erroneous information
      - Needs strict policies about who enters in data
      - Those in database should be able to view the data to check for errors and make sure personal information is correct and/or changed as needed
      - Old data no longer needed needs to be removed
      - Regular monitoring of data done either by human or software technology to weed out dirty data
- Storage of data
  - Data needs to be stored in multiple locations and regularly backed up
  - Backups should be made daily and weekly
  - Data storage area’s temperature needs to be monitored so the computer equipment doesn’t get too hot
  - Data storage area needs to be in a place safe from environmental interference (hurricanes, floods, etc...)
  - Copy of back-up data needs to be kept off-site or in a secure fire/flood proof safe
  - Regular rotation of back-up media
  - Back-up data needs to be checked for reliability (correct, current and complete information)
  - Backups need to be kept secure from unauthorized access